

# PERFORMANCE ANALYSIS AND OPTIMIZATION OF A CYCLIC-SERVICE TANDEM QUEUEING SYSTEM WITH MULTI-CLASS CUSTOMERS

Tsuyoshi Katayama

Department of Electronics and Informatics, Faculty of Engineering  
Toyama Prefectural University, Toyama 939-03, Japan

**Abstract**—This paper analyzes a special single-server tandem queueing system with feedback, which appears in call processing for multi-class calls (tasks) in telecommunication systems. Explicit expressions are derived for joint queue-length generating functions and the mean total sojourn time spent by a call (task) in the tandem queueing system. Using the interchange argument and the extended Kleinrock's conservation law, an optimal task scheduling strategy (called as the  $K/h$  rule) is obtained to minimize a cost function defined by the individual mean total sojourn times of multi-class tasks.

## 1. INTRODUCTION

The Integrated Services Digital Network (ISDN) has been extensively developed and introduced. It enables efficient transmission of various traffic demands. However, owing to multiple grades of service requirements, it needs essentially an optimal call processing for multi-class calls in the network nodes. We consider priority assignment problems on the call processing with multi-class tasks in switching systems. We use a single-server two-stage tandem queueing model with parallel queues for multi-class tasks in the second stage.

There have been analytical studies of tandem queues with cyclic-service. Nair [1] and Netto [2] considered a two-stage tandem queues with several switching rules, e.g., exhaustive, gated and limited services. The switching rules correspond to the task-scheduling for call processing in switching systems. Various switching rules of single-server queueing network systems have been studied in several literature, where these are different from this paper in decision epochs, i.e., *switching epochs*, (see Section 2 and [3-5]). By formulating as a Markov decision process, Browne and Yechiali [6], Yoshida *et al.* [7] and Liu and Nain [8] have recently studied priority assignment problems for the cyclic-service polling systems which are similar to this paper. Boxma *et al.* [9,10] have also investigated an optimization of polling systems by using the pseudo-conservation law. Optimization in cyclic-service tandem queueing systems is a subject which has so far received very little attention in queueing literature in [10] and [11].

In this paper, we would like to obtain explicit expressions of individual performance measures (waiting time, total sojourn time for each class task, etc.) and to find such a task-scheduling strategy for multi-class tasks as the  $K/h$  rule (also called the  $c\mu$  rule) in the cyclic-service tandem queueing system.

## 2. CYCLIC-SERVICE TANDEM QUEUEING MODEL

The queueing system considered below is a single-server two-stage tandem queue with feedback. The first stage has a common queue,  $Q_0$ , and the second stage has  $N$  parallel queues,  $Q_1, Q_2, \dots$  and  $Q_N$  for multi-class customers. Each queue,  $Q_n$ , has a service counter,  $S_n$ ,  $n = 1, 2, \dots, N$ . Customers of type- $n$  (or class- $n$ ) arrive at the common queue  $Q_0$  according to a Poisson process with rate  $\lambda_n$ ,  $n = 1, 2, \dots, N$ . After receiving the service in  $S_0$ , each customer either joins the end of a waiting line in  $Q_0$ , again with probability  $p$ , or departs from the first stage with probability  $q = 1 - p$ . Type- $n$  customers who have completed service in the first stage go immediately to  $Q_n$  to receive the second service in  $S_n$ ,  $n = 1, 2, \dots, N$ . All the queues are served by a single server

Typeset by  $\text{\texttt{AMS-TEX}}$

according to the exhaustive service discipline (i.e., the server continues to serve until each queue becomes empty) in cyclic order ( $S_0 \rightarrow S_1 \rightarrow S_2 \rightarrow \dots \rightarrow S_N \rightarrow S_0 \rightarrow S_1 \rightarrow$  and so on). We denote the switching rule in the second stage by  $P := \{S_1, S_2, \dots, S_N\}$ . Customers are served in each queue on the first-in-first-out (FIFO) discipline. Service times,  $\tau_n$ , at each counter  $S_n$ ,  $n = 0, 1, 2, \dots, N$  have a general distribution function  $H_n(t)$ , with finite first and second moments,  $h_n$  and  $h_n^{(2)}$ . The Laplace-Stieltjes transform (LST) of  $H_n(t)$  is denoted by  $H_n^*(s)$ ,  $n = 0, 1, 2, \dots, N$ . The switchover time (see, e.g., [12]) is assumed to be zero. (Since  $\tau_0 \ll \tau_n$ ,  $n = 1, 2, \dots, N$  in real switching systems, the first stage service corresponds to "pre-processing" and the second stage service to "main-processing" for multi-class tasks). The following notation is introduced:

$$\lambda := \sum_{i=1}^N \lambda_i \quad p_n := \frac{\lambda_n}{\lambda} \quad \rho_0 := \lambda h_0 \quad \rho_n := \lambda_n h_n \quad n = 1, 2, \dots, N. \quad (2.1)$$

Denoting by  $H(t)$  the distribution function of the total service time for a customer in the first stage, and using  $H^*(s)$  to denote the LST of  $H(t)$ , we have

$$H(t) := \sum_{k=1}^{\infty} H_0^{(k)}(t) (qp^{k-1}) \quad H^*(s) = \frac{qH_0^*(s)}{1 - pH_0^*(s)} \quad (2.2)$$

where  $H_0^{(k)}(t)$  denotes the  $k^{\text{th}}$  iterated convolution of  $H_0(t)$  with itself (see, e.g., Takács [13]). Let  $\alpha$  and  $\alpha^{(2)}$  denote the mean and the second moment for the distribution  $H(t)$ , then we get

$$\lambda := \frac{h_0}{q} \quad \alpha^{(2)} := \frac{h_0^{(2)}}{q} + \frac{2ph_0^2}{q^2}. \quad (2.3)$$

Thus, we define total server utilization:

$$\eta := \eta_0 + \sum_{n=1}^N \rho_n \quad \eta_0 := \lambda\alpha. \quad (2.4)$$

The server utilization  $\eta < 1$  is assumed for stability.

We denote by  $q_i(m)$  the probability that  $m$  customers arrive at  $Q_0$  during service times,  $\tau_n$ ,  $n = 0, 1, 2, \dots, N$ , i.e.,

$$q_n(m) := \int_0^{\infty} \frac{(\lambda t)^m}{m!} e^{-\lambda t} dH_n(t). \quad (2.5)$$

The generating function for  $q_n(m)$ ,  $Q_n(x) := \sum_{m=0}^{\infty} q_n(m) x^m$  is then given by

$$Q_n(x) = H_n^*\{\lambda(1-x)\} \quad n = 0, 1, 2, \dots, N. \quad (2.6)$$

We define similarly,

$$R_0(x) := \sum_{m=0}^{\infty} r_0(m) x^m = H^*\{\lambda(1-x)\} \quad (2.7)$$

where

$$r_0(m) := \int_0^{\infty} \frac{(\lambda t)^m}{m!} e^{-\lambda t} dH(t). \quad (2.8)$$

The following notation is used for a differentiable generating function  $G_n(x, y)$ ,

$$G'_{n,x}(a, b) := \left[ \left( \frac{\partial}{\partial x} \right) G_n(x, y) \right]_{x=a, y=b}. \quad (2.9)$$

## 3. QUEUEING ANALYSIS

This section determines a generating function of the following joint queue-length distribution at departure epochs of customers from each stage, which are defined by:

$\pi_n(i; j_1, j_2, \dots, j_N)$ : the steady-state joint probability that the number of waiting customers in  $Q_0, Q_1, Q_2, \dots$ , and  $Q_N$ , is  $i, j_1, j_2, \dots$ , and  $j_N$ , respectively, just after a customer has completed service at  $S_n$ ,  $n = 0, 1, 2, \dots, N$  and for  $x/y_k \leq 1, k = 1, 2, \dots, N$ ,

$$G_n(x; y_1, y_2, \dots, y_N) := \sum_{i=0}^{\infty} \sum_{j_1=0}^{\infty} \cdots \sum_{j_N=0}^{\infty} \pi_n(i; j_1, j_2, \dots, j_N) x^i y_1^{j_1} y_2^{j_2} \cdots y_N^{j_N}. \quad (3.1)$$

Then, considering the events that occur during two successive departure epochs at the first stage ( $S_0$ ), we have the following balance equations for  $i, j_1, \dots, j_N \geq 0$ ;

$$\begin{aligned} \pi_0(i; 0, \dots, 0, j_k = 1, 0, \dots, 0) &= \sum_{n=1}^N \pi_n(0; 0, \dots, 0) p_k r_0(i) \\ &+ \sum_{n=1}^N \sum_{m=0}^i \pi_n(i-m+1; 0, \dots, 0) p_k r_0(m) \quad k = 1, 2, \dots, N \end{aligned} \quad (3.2)$$

$$\pi_0(i; j_1, j_2, \dots, j_N) = \sum_{n=1}^N \sum_{m=0}^i \pi_0(i-m+1; j_1, \dots, j_{n-1}, j_n-1, j_{n+1}, \dots, j_N) p_n r_0(m),$$

where we use a relationship as in [13] that the customer departure process in the first stage is the same as the one in the first stage ( $S_0$ ) without feedback having arrival rate  $\lambda$  and the service time distribution  $H(t)$ . Similarly, considering the customer departure events from the service counter  $S_n$ ,  $n = 1, 2, \dots, N$ , in the second stage, we get the following equation for  $i, j_1, \dots, j_N \geq 0$ :

$$\begin{aligned} \pi_n(i; 0, \dots, 0, j_n, \dots, j_N) &= \pi_0(0; 0, \dots, 0, j_n+1, j_{n+1}, \dots, j_N) q_n(i) \\ &+ \sum_{m=0}^i \pi_n(i-m; 0, \dots, 0, j_n+1, j_{n+1}, \dots, j_N) q_n(m) \\ &+ \sum_{k=1}^{n-1} \sum_{m=0}^i \pi_k(i-m; 0, \dots, 0, j_n+1, j_{n+1}, \dots, j_N) q_n(m), \end{aligned} \quad (3.3)$$

$n = 1, 2, \dots, N,$

where the empty sum is assumed to be zero. Then, from (3.1) to (3.3), we obtain the following functional relations through routine calculations:

$$\begin{aligned} G_0(x; y_1, y_2, \dots, y_N) &= \frac{\sum_{n=1}^N p_n y_n R_0(x)}{x - \sum_{n=1}^N p_n y_n R_0(x)} \\ &\times [-(1-x) \pi(0) + \Phi\{Q_1(x), Q_2(x), \dots, Q_N(x)\} - \Phi\{y_1, y_2, \dots, y_N\}] \end{aligned} \quad (3.4a)$$

$$\begin{aligned} G_n(x; 0, \dots, 0, y_n, y_{n+1}, \dots, y_N) &= \frac{Q_n(x)}{y_n - Q_n(x)} \\ &\times [\Phi\{Q_1(x), Q_2(x), \dots, Q_{n-1}(x), y_n, \dots, y_N\} - \Phi\{Q_1(x), Q_2(x), \dots, Q_n(x), y_{n+1}, \dots, y_N\}], \end{aligned} \quad (3.4b)$$

$n = 1, 2, \dots, N,$

where

$$\begin{aligned} \pi(0) &:= \sum_{n=1}^N \pi_n(0; 0, \dots, 0) \\ \Phi\{y_1, y_2, \dots, y_N\} &:= G_0(0; y_1, y_2, \dots, y_N). \end{aligned} \quad (3.4c)$$

We see that it is necessary to determine unknown probability  $\pi(0)$  and unknown function  $\Phi\{y_1, y_2, \dots, y_N\}$ .

(1) Determination of  $\pi(0)$

From the normalization condition

$$G_0(1; 1, \dots, 1) + \sum_{n=1}^N G_n(1; 1, \dots, 1) = 1 \quad (3.5)$$

and using L'Hospital's rule for (3.4a) and (3.4b), we have

$$\pi(0) = \frac{1}{2} (1 - \eta) \quad (3.6)$$

and

$$G_0(1; 1, \dots, 1) = \frac{1}{2} \quad (3.7)$$

$$G_n(1; 0, \dots, 0, y_n = 1, 1, \dots, 1) = \frac{p_n}{2} \quad n = 1, 2, \dots, N.$$

(2) Determination of  $\Phi\{y_1, y_2, \dots, y_N\}$

It is shown by Takács' lemma [14] that the denominator on the right-hand side of (3.4),

$$x - \omega H^*\{\lambda(1 - x)\} = 0 \quad \omega := \sum_{n=1}^N p_n y_n \quad (3.8)$$

has exactly one root, namely  $x = \delta(\omega)$ , in the unit circle  $|x| \leq 1$  under the condition that  $\omega = 1$ ,  $\eta_0 \leq 1$  and  $|\omega| < 1$ . (The explicit expression for  $\delta(\omega)$  is given by Appendix in [14], for example.) Since the numerator on the right-hand side of (3.4a) should be zero for  $x = \delta(\omega)$ , we obtain the following linear functional equation

$$\Phi\{Q_1(\delta), Q_2(\delta), \dots, Q_N(\delta)\} - \Phi\{y_1, y_2, \dots, y_N\} = \phi\{y_1, y_2, \dots, y_N\} \quad (3.9)$$

where

$$\delta := \delta \left( \sum_{n=1}^N p_n y_n \right) \quad (3.10)$$

$$\phi(y_1, y_2, \dots, y_N) := \frac{1}{2} (1 - \eta) (1 - \delta)$$

This functional equation for  $\Phi\{\cdot\}$  can be solved using an iterative scheme and a boundary condition  $\Phi\{0, 0, \dots, 0\} \equiv G_0(0; 0, \dots, 0) = 0$  (see [15]). In this way, the generating function  $G_n(x; y_1, y_2, \dots, y_N)$ ,  $n = 0, 1, 2, \dots, N$  have been completely determined. Thus, we obtain the following results.

**THEOREM 3.1.** *The generating functions,  $G_n(x; y_1, y_2, \dots, y_N)$ ,  $n = 0, 1, 2, \dots, N$  for the joint queue-length distribution are given by (3.4a) and (3.4b), here*

$$\Phi\{y_1, y_2, \dots, y_N\} = \frac{1 - \eta}{2} \sum_{m=0}^{\infty} \left[ \delta \left\{ \sum_{n=1}^N p_n f_n^{(m)}(y_1, y_2, \dots, y_N) \right\} - \delta \left\{ \sum_{n=1}^N p_n f_n^{(m)}(0; 0, \dots, 0) \right\} \right] \quad (3.11a)$$

where,  $f_n^{(m)}(y_1, y_2, \dots, y_N)$  are defined by:

$$f_n^{(0)}(y_1, y_2, \dots, y_N) := y_n \quad |y_n| \leq 1 \quad (3.11b)$$

$$f_n^{(m+1)}(y_1, y_2, \dots, y_N) := Q_n \left[ \delta \left\{ \sum_{j=1}^N p_j f_j^{(m)}(y_1, y_2, \dots, y_N) \right\} \right] \quad m = 0, 1, 2, \dots; n = 1, 2, \dots, N.$$

#### 4. ANALYSIS OF MEAN SOJOURN TIMES

This section derives some explicit expressions for mean sojourn time of a type- $n$  customer,  $n = 1, 2, \dots, N$ . Let  $\theta_1^{(n)}$ ,  $n = 1, 2, \dots, N$  denote the sojourn time for a customer of type- $n$  in the first stage (the total time spent by a customer of type- $n$  in the first stage) and denote by  $w_1$  the waiting time in  $Q_0$  for an arbitrary customer including cycled customers. Similarly, denote by  $\theta_2^{(n)}$  and  $w_2^{(n)}$ ,  $n = 1, 2, \dots, N$  the sojourn time and the waiting time for a customer of type- $n$  in the second stage. In addition, denote by  $\theta^{(n)}$  and  $\theta_w^{(n)}$  the total sojourn time spent by a customer of type- $n$  in the system and the total waiting time for the second service of type- $n$  customer, respectively, i.e.,

$$\theta^{(n)} := \theta_1^{(n)} + \theta_2^{(n)} \quad \theta_w^{(n)} := \theta_1^{(n)} + \theta_2^{(n)} \quad n = 1, 2, \dots, N. \quad (4.1)$$

##### 4.1. Mean Sojourn Time in the First Stage

We first derive the mean queue-length,  $E(L_o)$ , (the mean number of customers waiting and being served) in the first stage at arbitrary epoch. Denoting by  $\Pi(x)$  the generating function for a marginal queue-length distribution in  $Q_0$  at the instant just after customer-departures from the first stage,  $\Pi(x)$  is given by

$$\Pi(x) = \frac{G_0(x; 1, 1, \dots, 1)}{G_0(1; 1, 1, \dots, 1)}. \quad (4.2)$$

Hence, from the property of the  $M/G/1$  type queue as in [16,17], the mean queue-length in  $Q_0$ ,  $E(L_o)$ , is also obtained by

$$E(L_o) = \frac{G'_{0,x}(1; 1, 1, \dots, 1)}{G_0(1; 1, 1, \dots, 1)}. \quad (4.3)$$

Next, denoting by  $E(Q_o)$  the expected number of waiting customer in  $Q_o$  at arbitrary epoch, we get

$$E(L_o) = E(Q_o) + 1 - \Pi(0). \quad (4.4)$$

Thus, using Little's formula, Equation (3.7) and the total arrival rate to  $Q_o (= \lambda/q)$ , we obtain the following results.

**THEOREM 4.1.** *The mean sojourn time in the first stage of a type- $n$  customer,  $E(\theta_1^{(n)})$ ,  $n = 1, 2, \dots, N$  and the mean waiting time in  $Q_o$ ,  $E(w_1)$ , are given by:*

$$E(\theta_1^{(n)}) = 2G'_{0,x}(1; 1, \dots, 1) \frac{1}{\lambda} \quad n = 1, 2, \dots, N \quad (4.5a)$$

$$E(w_1) = \frac{2q}{\lambda} \left\{ G'_{0,x}(1; 1, \dots, 1) + G_0(o; 1, \dots, 1) - \frac{1}{2} \right\}. \quad (4.5b)$$

##### 4.2. Mean Sojourn Time in the Second Stage

We first obtain the condition expectation for sojourn time distribution of a type- $n$  customer in the second stage,  $E(\theta_2^{(n)} | \mathbf{J})$ , where  $\mathbf{J} := \{i; j_1, \dots, j_n + 1, \dots, j_N\}$  represents the system state that the number of waiting customer in  $Q_0, Q_1, \dots, Q_n, \dots$  and  $Q_N$  is  $i, j_1, \dots, j_n + 1, \dots$  and  $j_N$  when a tagged customer of type- $n$ ,  $C_n^*$  has arrived at the second queue,  $Q_n$ ,  $n = 1, 2, \dots, N$ . Since the expected number of customers served at  $S_0$  during the  $i$ -busy period is equal to  $i/(1 - \eta_0)$ , we get

$$E(\theta_2^{(n)} | \mathbf{J}) = \frac{i\alpha}{1 - \eta_0} + \sum_{k=1}^{n-1} \left\{ \left( j_k + \frac{ip_k}{1 - \eta_0} \right) h_k \right\} + (j_n + 1) h_n. \quad (4.6)$$

Next, we need to determine a conditional probability denoted by  $\pi_0(i; j_1, \dots, j_n, \dots, j_N | n)$ , that queue-length in  $Q_0, Q_1, \dots, Q_n, \dots$  and  $Q_N$  is  $i, j_1, \dots, j_n (\geq 1), \dots$  and  $j_N$ , when the tagged customer  $C_n^*$  has arrived at the queue,  $Q_n$ ,  $n = 1, 2, \dots, N$ . In order to facilitate the determination of the generating function  $G_0(x; y_1, y_2, \dots, y_N | n)$  for the distribution

$\{\pi_0(i; j_1, \dots, j_N \mid n)\}$ , we introduce the probability denoted by  $\pi_0^*(i; j_1, \dots, j_N)$  that at the instant just before customer-departures from the first stage, queue-length in  $Q_0, Q_1, \dots$  and  $Q_N$  is  $i(\geq 1), j_1, \dots$  and  $j_N$ , respectively. Denoting by  $G_0^*(x; y_1, y_2, \dots, y_N)$  the generating function for the distribution  $\{\pi_0^*(i; j_1, \dots, j_N)\}$ , we have

$$\frac{G_0(x; y_1, y_2, \dots, y_N)}{G_0(1; 1, 1, \dots, 1)} = G_0^*(x; y_1, y_2, \dots, y_N) \frac{\sum_{n=1}^N p_n y_n}{x} \quad (4.7a)$$

$$G_0(x; y_1, y_2, \dots, y_N \mid n) = G_0^*(x; y_1, y_2, \dots, y_N) \cdot \frac{y_n}{x}. \quad (4.7b)$$

Hence, using (4.7a), (4.7b), and (3.7), we get

$$\begin{aligned} G_0(x; y_1, y_2, \dots, y_N \mid n) &= \frac{G_0(x; y_1, y_2, \dots, y_N)}{G_0(1; 1, 1, \dots, 1)} \left\{ \frac{y_n}{\sum_{n=1}^N p_n y_n} \right\} \\ &= \frac{2y_n}{\sum_{n=1}^N p_n y_n} G_0(x; y_1, y_2, \dots, y_N). \end{aligned} \quad (4.8)$$

Thus, removing the condition **J** from (4.6) by using (4.8), we get the following results.

**THEOREM 4.2.** *The mean sojourn time in the second stage of a type- $n$  customer,  $E(\theta_2^{(n)})$ , and the mean waiting time in the second queue,  $E(w_2^{(n)})$ ,  $n = 1, 2, \dots, N$  are given by*

$$\begin{aligned} E(\theta_2^{(n)}) &= \sum_{i=0}^{\infty} \sum_{j_1=0}^{\infty} \cdots \sum_{j_N=0}^{\infty} E(\theta_2^{(n)} \mid \mathbf{J}) \pi_0(i; j_1, \dots, j_n + 1, \dots, j_N \mid n) \\ &= h_n - \sum_{k=1}^n p_k h_k + \frac{2}{1 - \eta_0} \left\{ \alpha + \sum_{k=1}^{n-1} p_k h_k \right\} G'_{0,x}(1; 1, \dots, 1) \end{aligned} \quad (4.9a)$$

$$\begin{aligned} &+ 2 \sum_{k=1}^n h_k G'_{0,y_k}(1; 1, \dots, 1) \\ E(w_2^{(n)}) &= E(\theta_2^{(n)}) - h_n \quad n = 1, 2, \dots, N. \end{aligned} \quad (4.9b)$$

The following inequality for  $E(w_2^{(n)})$  is derived as being naturally expected:

**COROLLARY 4.1.** *The switching rule,  $P = \{S_1, S_2, \dots, S_N\}$  implies*

$$E(w_2^{(1)}) < E(w_2^{(2)}) < \cdots < E(w_2^{(N)}). \quad (4.10)$$

**PROOF.** From (4.7a) and (4.7b), we get

$$\begin{aligned} \pi_0^*(i+1; j_1, \dots, j_n, \dots, j_N) &= \pi_0(i; j_1, \dots, j_{n-1}, j_n + 1, j_{n+1}, \dots, j_N \mid n) \\ &= \pi_0(i; j_1, \dots, j_n, j_{n+1} + 1, j_{n+2}, \dots, j_N \mid n+1). \end{aligned} \quad (4.11)$$

Using (4.6), (4.9a), (4.9b) and (4.11), the difference ( $D$ ) between  $E(w_2^{(n+1)})$  and  $E(w_2^{(n)})$  is expressed by

$$D = \sum_{i=0}^{\infty} \sum_{j_1=0}^{\infty} \cdots \sum_{j_N=0}^{\infty} \pi_0^*(i+1; j_1, \dots, j_n, \dots, j_N) \left\{ \left( \frac{ip_n}{1 - \eta_0} + j_n \right) h_n + j_{n+1} h_{n+1} - j_n h_n \right\}. \quad (4.12)$$

Since the right-hand side of (4.12) is positive ( $D > 0$ ), we get (4.10).

#### 4.9. Mean Total Sojourn Time

The mean total sojourn time and the mean total waiting time of a type- $n$  customer are given by

$$\begin{aligned} E(\theta^{(n)}) &= E(\theta_1^{(n)}) + E(\theta_2^{(n)}) \\ E(\theta_w^{(n)}) &= E(\theta_1^{(n)}) + E(w_2^{(n)}) \quad n = 1, 2, \dots, N. \end{aligned} \quad (4.13)$$

Hence, Corollary 4.1 is rewritten as follows:

**COROLLARY 4.2.** *The switching rule,  $P = \{S_1, S_2, \dots, S_N\}$  implies*

$$E(\theta_w^{(1)}) < E(\theta_w^{(2)}) < \dots < E(\theta_w^{(N)}). \quad (4.14)$$

**REMARK 4.1.** Note that the inequality (4.14) does not always hold for  $E(\theta^{(n)})$ ,  $n = 1, 2, \dots, N$ . It can be shown in a similar manner to the proof of Corollary 4.1.

### 5. OPTIMIZATION OF THE SWITCHING RULE IN THE SECOND STAGE

In this section, we investigate an optimal switching rule in the second stage so as to minimize a total sojourn time cost defined by

$$C := \sum_{n=1}^N p_n \{K_n E(\theta^{(n)})\} \quad (5.1)$$

where  $K_n$  is the delay cost for a customer of type- $n$ .

#### 5.1. Conservation Law

To find the optimal switching rule in the second stage, the following lemma will be necessary.

**LEMMA 5.1.** *For the two-stage tandem queue with feed-back defined in Section 2, the value of  $V$  defined by (5.2) is invariant for the work-conserving switching rule:*

$$V := \frac{1}{q} \left( \eta_0 + \sum_{n=1}^N \rho_n \right) E(w_1) + \sum_{n=1}^N \rho_n E(w_2^{(n)}). \quad (5.2)$$

**PROOF.** The total arrival rate to  $Q_0$  is equal to  $\lambda/q$  and the mean total service time in the system to  $\{\alpha + \sum_{n=1}^N p_n h_n\}$ . Thus, the relationship (5.2) is directly derived from Theorem 4.2 in [18].

(This lemma is an extension of the conservation law first given by Kleinrock and can be also obtained from Theorem 3.1 and Remark 3.2 in [19].)

#### 5.2. Optimal Switching Rule in the Second Stage

We shall derive the following theorem, known as the  $K/h$  rule in [20], by using the interchange argument and Lemma 5.1.

**THEOREM 5.1.** *The optimal priority assignment to minimize the total sojourn time cost,  $C$ , is in descending order of  $K_n/h_n$ , with the highest priority for the type with the highest value of  $K_n/h_n$ ,  $n = 1, 2, \dots, N$ .*

**PROOF.** Consider a switching rule in which the priority of just two neighboring types, say  $j$  and  $j+1$  are interchanged, and denote the switching rules before and after the interchange as  $P = \{S_1, S_2, \dots, S_N\}$  and  $P^\#$ , that is,

$$P^\# := \{S_1, S_2, \dots, S_{j+1}, S_j, \dots, S_N\}. \quad (5.3)$$

For all relevant queueing quantities under the condition of  $P^\#$ , the same notation modified by adding the symbol  $\#$ , for example,  $E(\theta^\#_{i^{(n)}})$ ,  $E(w^\#_{i^{(n)}})$ ,  $i = 1, 2$ ,  $\pi^\#_0(i; j_1, j_2, \dots, j_N \mid n)$ ,

$n = 1, 2, \dots, N$  and so on is used. First, we compare two mean total sojourn times for a customer of type- $n$  under the switching rules,  $P$  and  $P^\#$ . Since the generating function  $\Pi(x)$  in (4.2) for the queue length distribution in  $Q_0$  is not affected by the interchange of  $S_j$  and  $S_{j+1}$ , we get the following equation by using Theorem 4.1:

$$E(\theta_1^{(n)}) = E(\theta^\#_1^{(n)}) \quad E(w_1) = E(w^\#_1). \quad (5.4)$$

Equations (4.8) and (4.9a) also yield

$$\pi_0(i : j_1, j_2, \dots, j_N | n) = \pi^\#_0(i : j_1, j_2, \dots, j_N | n) \quad (5.5)$$

$$E(\theta_2^{(n)}) = E(\theta^\#_2^{(n)}) \quad \text{for } n < j \text{ and } n > j + 1. \quad (5.6)$$

Lemma 5.1 and the last two equalities, (5.4) and (5.6), imply

$$\rho_j E(w_2^{(j)}) + \rho_{j+1} E(w_2^{(j+1)}) = \rho_j E(w^\#_2^{(j)}) + \rho_{j+1} E(w^\#_2^{(j+1)}). \quad (5.7)$$

Thus, the difference ( $\Delta$ ) of two total sojourn time costs for the switching rules,  $P$  and  $P^\#$  is expressed by

$$\begin{aligned} \Delta &:= C - C^\# \\ &= \rho_j \rho_{j+1} \{E(w_2^{(j)}) - E(w^\#_2^{(j)})\} \left\{ \frac{K_j}{h_j} - \frac{K_{j+1}}{h_{j+1}} \right\}. \end{aligned} \quad (5.8)$$

Using (5.5), we need to evaluate the following difference,

$$\begin{aligned} E(w_2^{(j)}) - E(w^\#_2^{(j)}) &= \sum_{i=0}^{\infty} \sum_{k_1=0}^{\infty} \cdots \sum_{k_N=0}^{\infty} \pi_0(i; k_1, \dots, k_{j-1}, k_j + 1, k_{j+1}, \dots, k_N | J) \\ &\quad \cdot \left[ h_j(k_j + 1) - \left\{ h_{j+1} \left( k_{j+1} + p_{j+1} \cdot \frac{i}{1 - \eta_0} \right) + h_j(k_j + 1) \right\} \right]. \end{aligned} \quad (5.9)$$

Hence, we get  $E(w_2^{(j)}) - E(w^\#_2^{(j)}) < 0$ . (This inequality is also clear from the property of the priority switching rule). Therefore,  $C^\# > C$ , or a type- $j$  customer should be given higher priority than a type- $(j + 1)$  customer if

$$\frac{K_j}{h_j} > \frac{K_{j+1}}{h_{j+1}}. \quad (5.10)$$

Thus, using (5.10), the optimal switching rule can be obtained by interchanging some neighboring classes successively. Hence, we get the optimal priority assignment.

Next, we consider the following cost function,

$$C^* := \sum_{n=1}^N K_n E(\theta^{(n)}). \quad (5.11)$$

Using  $K_n$  instead of  $p_n K_n$  in (5.1) and the proof of Theorem 5.1, Theorem 5.1 can be expressed as follows:

**COROLLARY 5.1.** *The optimal priority assignment to minimize the total waiting time cost,  $C^*$ , is in descending order of  $K_n/\rho_n$ , with the highest priority for the type with the highest value of  $K_n/\rho_n$ .*

**REMARK 5.1.** It is also shown that Theorem 5.1 and Corollary 5.1 hold for the cases of total waiting time costs,

$$C_w := \sum_{n=1}^N p_n \{K_n E(\theta_w^{(n)})\} \quad (5.12a)$$

and

$$C_w^* := \sum_{n=1}^N K_n E(\theta_w^{(n)}), \quad (5.12b)$$

respectively (see, e.g., [20]).



## 6. CONCLUSIONS

The study of priority assignments of multi-class calls is important to design various call processing programs in multi-media network nodes. From theoretical points of view, we analyzed the basic cyclic-service tandem queueing system with multi-class tasks. Some explicit expressions were obtained for individual performance measures, i.e., the mean total sojourn time of a type- $n$  task,  $n = 1, 2, \dots, N$ . Using the results and the well-known interchange argument, the optimal task scheduling known as the  $K/h$  rule (also called the  $c\mu$  rule) was derived.

Further research will be extended to the evaluation of the influence caused by the switch over time (also called walking time and overhead time) on some performance measures, and to the generalization of arrival processes, e.g., non-Poissonian renewal arrivals, correlated arrivals, batch arrivals and so on.

## REFERENCES

1. S.S. Nair, Semi-Markov analysis of two queues in series attended by a single server, *Bull. Soc. Math. Belgique* **22**, 355–367 (1971).
2. M.T. Netto, Two queues in tandem attended by a single server, *Oper. Res.* **25** (1), 140–147 (1977).
3. M. Hofri and K.W. Ross, On the optimal control of two queues with server setup times and its analysis, *SIAM J. Comput.* **16** (2), 399–420 (1987).
4. I. Meilijson and G. Weiss, Multiple feedback at a single-server station, *Stochastic Processes and Their Application* **5** (2), 195–205 (1977).
5. D. Tcha and S.R. Pliska, Optimal control of single server queueing networks, *Oper. Res.* **25** (2), 248–258 (1977).
6. S. Browne and U. Yechiali, Dynamic priority rules for cyclic-type queues, *Adv. Appl. Prob.* **21** (3), 432–450 (1989).
7. H. Yoshida, M. Ohnishi and T. Ibaraki, Optimal visiting policies in symmetric polling systems, In *Abstracts of the 1990 Spring National Conference of ORSJ*, 2-A-9, pp. 144–145, (1990).
8. Z. Liu and P. Nain, Optimal scheduling in some multi-queue single-server systems, In *Proceedings of INFOCOM '90*, pp. 1213–1219, (1990).
9. O.J. Boxma, H. Levy and J.A. Weststrate, Optimization of polling systems, In *Performance '90*, (Edited by P.J.B. King, I. Mittrani and R.J. Pooley), pp. 349–361, North-Holland, Amsterdam, (1990).
10. O.J. Boxma, Analysis and optimization of polling systems, In *Queueing, Performance and Control in ATM (ITC-13)*, (Edited by J.W. Cohen and C.D. Pack), pp. 173–183, North-Holland, (1991).
11. U. Yechiali, Optimal dynamic control of polling systems, In *Queueing, Performance and Control in ATM (ITC-13)*, (Edited by J.W. Cohen and C.D. Pack), pp. 205–217, North-Holland, (1991).
12. H. Takagi, *Analysis of Polling Systems*, MIT Press, Cambridge networks, (1986).
13. L. Takács, A single-server queue with feedback, *Bell Syst. Tech. J.* **42** (2), 509–519 (1963).
14. J.W. Cohen, *The Single Server Queue*, (Revised edition), North-Holland, Amsterdam, (1982).
15. M. Kuczma, *Functional Equations in a Single Variable*, PWN-Polish Sci. Pub., Warszawa, (1968).
16. R.B. Cooper, *Introduction to Queueing Theory*, (Second edition), Edward Arnold, London, (1981).
17. R.W. Wolff, Poisson arrivals see time averages, *Oper. Res.* **30** (1), 223–231 (1982).
18. T. Katayama, A cyclic-service tandem queueing model with semi-exhaustive service, In *Performance of Distributed and Parallel Systems*, (Edited by T. Hasegawa, H. Takagi and Y. Takahashi), pp. 183–196, North-Holland, Amsterdam, (1989).
19. O.J. Boxma, Workloads and waiting times in single-server systems with multiple customer classes, *Queueing Systems* **1–3** (5), 185–214 (1989).
20. E. Gelenbe and I. Mittrani, *Analysis and Synthesis of Computer Systems*, Academic Press, New York-London, (1980).
21. T. Katayama, Mean sojourn times in a multi-stage tandem queue served by a single server, *J. Oper. Res. Soc. Japan* **31** (2), 233–250 (1988).
22. L. Schrage, An alternative proof of a conservation law for the queue  $G/G/1$ , *Oper. Res.* **18** (1), 185–187 (1970).
23. M. Sidi and H. Levy, Customers routing in polling systems, In *Performance '90*, (Edited by P.J.B. King, I. Mittrani and R.J. Pooley), North-Holland, Amsterdam, pp. 319–331, (1990).